

COPY OF  
ORIGINAL

RECEIVED

SEQUENCE LISTING

MAR - 1 2002

TECH CENTER 1600/2900

## (1) GENERAL INFORMATION

- (i) APPLICANT: Lal, Preeti  
Bandman, Olga
- (ii) TITLE OF THE INVENTION: NOVEL HUMAN SODIUM-DEPENDENT  
PHOSPHATE COTRANSPORTER
- (iii) NUMBER OF SEQUENCES: 7
- (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: Incyte Pharmaceuticals, Inc.  
(B) STREET: 3174 Porter Drive  
(C) CITY: Palo Alto  
(D) STATE: CA  
(E) COUNTRY: US  
(F) ZIP: 94304
- (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Diskette  
(B) COMPUTER: IBM Compatible  
(C) OPERATING SYSTEM: DOS  
(D) SOFTWARE: FastSEQ Version 2.0
- (vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER: To Be Assigned  
(B) FILING DATE: Filed Herewith  
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER:  
(B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: Billings, Lucy J.  
(B) REGISTRATION NUMBER: 36,749  
(C) REFERENCE/DOCKET NUMBER: PF-0221 US
- (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: 415-855-0555  
(B) TELEFAX: 415-845-4166

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 401 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:  
(A) LIBRARY: BRAITUT02  
(B) CLONE: 754412

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Gln Val Asp Glu Thr Leu Ile Pro Arg Lys Val Pro Ser Leu Cys  
 1 5 10 15  
 Ser Ala Arg Tyr Gly Ile Ala Leu Val Leu His Phe Cys Asn Phe Thr  
 20 25 30  
 Thr Ile Ala Gln Asn Val Ile Met Asn Ile Thr Met Val Ala Met Val  
 35 40 45  
 Asn Ser Thr Ser Pro Gln Ser Gln Leu Asn Asp Ser Ser Glu Val Leu  
 50 55 60  
 Pro Val Asp Ser Phe Gly Gly Leu Ser Lys Ala Pro Lys Ser Leu Pro  
 65 70 75 80  
 Ala Lys Ser Ser Ile Leu Gly Gly Gln Phe Ala Ile Trp Glu Arg Trp  
 85 90 95  
 Gly Pro Pro Gln Glu Arg Ser Arg Leu Cys Ser Ile Ala Leu Ser Gly  
 100 105 110  
 Met Leu Leu Gly Cys Phe Thr Ala Ile Leu Ile Gly Gly Phe Ile Ser  
 115 120 125  
 Glu Thr Leu Gly Trp Pro Phe Val Phe Tyr Ile Phe Gly Gly Val Gly  
 130 135 140  
 Cys Val Cys Cys Leu Leu Trp Phe Val Val Ile Tyr Asp Asp Pro Val  
 145 150 155 160  
 Ser Tyr Pro Trp Ile Ser Thr Ser Glu Lys Glu Tyr Ile Ile Ser Ser  
 165 170 175  
 Leu Lys Gln Gln Val Gly Ser Ser Lys Gln Pro Leu Pro Ile Lys Ala  
 180 185 190  
 Met Leu Arg Ser Leu Pro Ile Trp Ser Ile Cys Leu Gly Cys Phe Ser  
 195 200 205  
 His Gln Trp Leu Val Ser Thr Met Val Val Tyr Ile Pro Thr Tyr Ile  
 210 215 220  
 Ser Ser Val Tyr His Val Asn Ile Arg Asp Asn Gly Leu Leu Ser Ala  
 225 230 235 240  
 Leu Pro Phe Ile Val Ala Trp Val Ile Gly Met Val Gly Gly Tyr Leu  
 245 250 255  
 Ala Asp Phe Leu Leu Thr Lys Lys Phe Arg Leu Ile Thr Val Arg Lys  
 260 265 270  
 Ile Ala Thr Ile Leu Gly Ser Leu Pro Ser Ser Ala Leu Ile Val Ser  
 275 280 285  
 Leu Pro Tyr Leu Asn Ser Gly Tyr Ile Thr Ala Thr Ala Leu Leu Thr  
 290 295 300  
 Leu Ser Cys Gly Leu Ser Thr Leu Cys Gln Ser Gly Ile Tyr Ile Asn  
 305 310 315 320  
 Val Leu Asp Ile Ala Pro Arg Tyr Ser Ser Phe Leu Met Gly Ala Ser  
 325 330 335  
 Arg Gly Phe Ser Ser Ile Ala Pro Val Ile Val Pro Thr Val Ser Gly  
 340 345 350  
 Phe Leu Leu Ser Gln Asp Pro Glu Phe Gly Trp Arg Asn Val Phe Phe  
 355 360 365  
 Leu Leu Phe Ala Val Asn Leu Leu Gly Leu Leu Phe Tyr Leu Ile Phe  
 370 375 380  
 Gly Glu Ala Asp Val Gln Glu Trp Ala Lys Glu Arg Lys Leu Thr Arg  
 385 390 395 400  
 Leu

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1643 base pairs

B1

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: BRAITUT02  
(B) CLONE: 754412

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

AGAACGGTGA GGATGACCGA CGTATAGGCG AGAGCCTAGG TACGCCATGC CAGGTCACCG 60
GTCCGGCAAT TCCCGGGTCG ACCCACGCGT CCGCTTGGAG GGACGCTGGG TTCAACTTGA 120
AGCCCTTCCA CAGACATTAA GTCGGTGAAA ACCATTCACT AGGAGAGGAG AAACACAATG 180
GCCACCAAGA CAGAGTTGAG TCCCACAGCA AGGGAGAGCA AGAACGCACA AGATATGCAA 240
GTGGATGAGA CACTGATCCC CAGGAAAGTT CCAAGTTTAT GTTCTGCTCG CTATGGAATA 300
GCCCTCGTCT TACATTTCTG CAATTTCAACA ACGATAGCAC AAAATGTCAT CATGAACATC 360
ACCATGGTAG CCATGGTCAA CAGCACAAAGC CCTCAATCCC AGCTCAATGA TTCCTCTGAG 420
GTGCTGCCTG TTGACTCATT TGGTGGCCTA AGTAAAGCCC CAAAGAGTCT TCCTGCAAAG 480
TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA 540
AGCAGACTCT GCAGCATTCG TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600
ATAGGTGGCT TCATTAGTGA AACCCCTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660
GTTGGCTGTG TCTGCTGCCT TCTCTGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720
CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780
TCTTCTAAGC AGCCTCTTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840
TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGTTGTATA CATACCAACT 900
TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960
TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CTTCTAACC 1020
AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCTCT 1080
TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCGGCT ATATCACAGC AACTGCCTTG 1140
CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200
GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TCGAGCATA 1260
GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320
TGGAGGAATG TCTTCTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380
ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACCTCAC TCGTTTATGA 1440
AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500
CCGTGATGAA AATACCACTG AAAAGATTTT TTTTCTCTGT GGCTCTTTTC AATTATGAGA 1560
TCAGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAATTCA 1620
AATAAAATGA TAACTAAGAA TGC 1643

```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 467 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:  
(A) LIBRARY: GenBank  
(B) CLONE: 450532

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Met Gln Met Asp Asn Arg Leu Pro Pro Lys Lys Val Pro Gly Phe Cys
1           5           10           15
Ser Phe Arg Tyr Gly Leu Ser Phe Leu Val His Cys Cys Asn Val Ile
20           25           30
Ile Thr Ala Gln Arg Ala Cys Leu Asn Leu Thr Met Val Val Met Val
35           40           45

```

31

Asn	Ser	Thr	Asp	Pro	His	Gly	Leu	Pro	Asn	Thr	Ser	Thr	Lys	Lys	Leu
50						55					60				
Leu	Asp	Asn	Ile	Lys	Asn	Pro	Met	Tyr	Asn	Trp	Ser	Pro	Asp	Ile	Gln
65					70					75					80
Gly	Ile	Ile	Leu	Ser	Ser	Thr	Ser	Tyr	Gly	Val	Ile	Ile	Ile	Gln	Val
				85					90					95	
Pro	Val	Gly	Tyr	Phe	Ser	Gly	Ile	Tyr	Ser	Thr	Lys	Lys	Met	Ile	Gly
			100					105					110		
Phe	Ala	Leu	Cys	Leu	Ser	Ser	Val	Leu	Ser	Leu	Leu	Ile	Pro	Pro	Ala
	115						120					125			
Ala	Gly	Ile	Gly	Val	Ala	Trp	Val	Val	Val	Cys	Arg	Ala	Val	Gln	Gly
130						135					140				
Ala	Ala	Gln	Gly	Ile	Val	Ala	Thr	Ala	Gln	Phe	Glu	Ile	Tyr	Val	Lys
145					150					155					160
Trp	Ala	Pro	Pro	Leu	Glu	Arg	Gly	Arg	Leu	Thr	Ser	Met	Ser	Thr	Ser
				165					170					175	
Gly	Phe	Leu	Leu	Gly	Pro	Phe	Ile	Val	Leu	Leu	Val	Thr	Gly	Val	Ile
		180						185					190		
Cys	Glu	Ser	Leu	Gly	Trp	Pro	Met	Val	Phe	Tyr	Ile	Phe	Gly	Ala	Cys
	195						200					205			
Gly	Cys	Ala	Val	Cys	Leu	Leu	Trp	Phe	Val	Leu	Phe	Tyr	Asp	Asp	Pro
210						215					220				
Lys	Asp	His	Pro	Cys	Ile	Ser	Ile	Ser	Glu	Lys	Glu	Tyr	Ile	Thr	Ser
225					230						235				240
Ser	Leu	Val	Gln	Gln	Val	Ser	Ser	Ser	Arg	Gln	Ser	Leu	Pro	Ile	Lys
				245					250					255	
Ala	Ile	Leu	Lys	Ser	Leu	Pro	Val	Trp	Ala	Ile	Ser	Ile	Gly	Ser	Phe
		260						265					270		
Thr	Phe	Phe	Trp	Ser	His	Asn	Ile	Met	Thr	Leu	Tyr	Thr	Pro	Met	Phe
	275						280					285			
Ile	Asn	Ser	Met	Leu	His	Val	Asn	Ile	Lys	Glu	Asn	Gly	Phe	Leu	Ser
290						295					300				
Ser	Leu	Pro	Tyr	Leu	Phe	Ala	Trp	Ile	Cys	Gly	Asn	Leu	Ala	Gly	Gln
305					310					315				320	
Leu	Ser	Asp	Phe	Phe	Leu	Thr	Arg	Asn	Ile	Leu	Ser	Val	Ile	Ala	Val
				325					330					335	
Arg	Lys	Leu	Phe	Thr	Ala	Ala	Gly	Phe	Leu	Leu	Pro	Ala	Ile	Phe	Gly
		340					345					350			
Val	Cys	Leu	Pro	Tyr	Leu	Ser	Ser	Thr	Phe	Tyr	Ser	Ile	Val	Ile	Phe
	355						360					365			
Leu	Ile	Leu	Ala	Gly	Ala	Thr	Gly	Ser	Phe	Cys	Leu	Gly	Gly	Val	Phe
	370					375					380				
Ile	Asn	Gly	Leu	Asp	Ile	Ala	Pro	Arg	Tyr	Phe	Gly	Phe	Ile	Lys	Ala
385					390					395				400	
Cys	Ser	Thr	Leu	Thr	Gly	Met	Ile	Gly	Gly	Leu	Ile	Ala	Ser	Thr	Leu
				405					410					415	
Thr	Gly	Leu	Ile	Leu	Lys	Gln	Asp	Pro	Glu	Ser	Ala	Trp	Phe	Lys	Thr
		420					425					430			
Phe	Ile	Leu	Met	Ala	Ala	Ile	Asn	Val	Thr	Gly	Leu	Ile	Phe	Tyr	Leu
	435						440					445			
Ile	Val	Ala	Thr	Ala	Glu	Ile	Gln	Asp	Trp	Ala	Lys	Glu	Lys	Gln	His
450						455					460				
Thr	Arg	Leu													
465															

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 560 amino acids

B1

(B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (vii) IMMEDIATE SOURCE:

(A) LIBRARY: GenBank  
 (B) CLONE: 507415

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Glu Phe Arg Gln Glu Glu Phe Arg Lys Leu Ala Gly Arg Ala Leu
 1           5           10          15
Gly Arg Leu His Arg Leu Leu Glu Lys Arg Gln Glu Gly Ala Glu Thr
          20           25           30
Leu Glu Leu Ser Ala Asp Gly Arg Pro Val Thr Thr His Thr Arg Asp
          35           40           45
Pro Pro Val Val Asp Cys Thr Cys Phe Gly Leu Pro Arg Arg Tyr Ile
          50           55           60
Ile Ala Ile Met Ser Gly Leu Gly Phe Cys Ile Ser Phe Gly Ile Arg
65           70           75           80
Cys Asn Leu Gly Val Ala Ile Val Ser Met Val Asn Asn Ser Thr Thr
          85           90           95
His Arg Gly Gly His Val Val Val Gln Lys Ala Gln Phe Asn Trp Asp
          100          105          110
Pro Glu Thr Val Gly Leu Ile His Gly Ser Phe Phe Trp Gly Tyr Ile
          115          120          125
Val Thr Gln Ile Pro Gly Gly Phe Ile Cys Gln Lys Phe Ala Ala Asn
          130          135          140
Arg Val Phe Gly Phe Ala Ile Val Ala Thr Ser Thr Leu Asn Met Leu
145          150          155          160
Ile Pro Ser Ala Ala Arg Val His Tyr Gly Cys Val Ile Phe Val Arg
          165          170          175
Ile Leu Gln Gly Leu Val Glu Gly Val Thr Tyr Pro Ala Cys His Gly
          180          185          190
Ile Trp Ser Lys Trp Ala Pro Pro Leu Glu Arg Ser Arg Leu Ala Thr
          195          200          205
Thr Ala Phe Cys Gly Ser Tyr Ala Gly Ala Val Val Ala Met Pro Leu
          210          215          220
Ala Gly Val Leu Val Gln Tyr Ser Gly Trp Ser Ser Val Phe Tyr Val
225          230          235          240
Tyr Gly Ser Phe Gly Ile Phe Trp Tyr Leu Phe Trp Leu Leu Val Ser
          245          250          255
Tyr Glu Ser Pro Ala Leu His Pro Ser Ile Ser Glu Glu Glu Arg Lys
          260          265          270
Tyr Ile Glu Asp Ala Ile Gly Glu Ser Ala Lys Leu Met Asn Pro Val
          275          280          285
Thr Lys Phe Asn Thr Pro Trp Arg Arg Phe Phe Thr Ser Met Pro Val
          290          295          300
Tyr Ala Ile Ile Val Ala Asn Phe Cys Arg Ser Trp Thr Phe Tyr Leu
305          310          315          320
Leu Leu Ile Ser Gln Pro Ala Tyr Phe Glu Glu Val Phe Gly Phe Glu
          325          330          335
Ile Ser Lys Val Gly Leu Val Ser Ala Leu Pro His Leu Val Met Thr
          340          345          350
Ile Ile Val Pro Ile Gly Gly Gln Ile Ala Asp Phe Leu Arg Ser Arg
          355          360          365
His Ile Met Ser Thr Thr Asn Val Arg Lys Leu Met Asn Cys Gly Gly
          370          375          380
Phe Gly Met Glu Ala Thr Leu Leu Leu Val Val Gly Tyr Ser His Ser

```

B1

385		390		395		400
Lys Gly Val Ala Ile Ser Phe Leu Val Leu Ala Val Gly Phe Ser Gly						
	405		410		415	
Phe Ala Ile Ser Gly Phe Asn Val Asn His Leu Asp Ile Ala Pro Arg						
	420		425		430	
Tyr Ala Ser Ile Leu Met Gly Ile Ser Asn Gly Val Gly Thr Leu Ser						
	435		440		445	
Gly Met Val Cys Pro Ile Ile Val Gly Ala Met Thr Lys His Lys Thr						
	450		455		460	
Arg Glu Glu Trp Gln Tyr Val Phe Leu Ile Ala Ser Leu Val His Tyr						
	465		470		475	
Gly Gly Val Ile Phe Tyr Gly Val Phe Ala Ser Gly Glu Lys Gln Pro						
	485		490		495	
Trp Ala Glu Pro Glu Glu Met Ser Glu Glu Lys Cys Gly Phe Val Gly						
	500		505		510	
His Asp Gln Leu Ala Gly Ser Asp Glu Ser Glu Met Glu Asp Glu Val						
	515		520		525	
Glu Pro Pro Gly Ala Pro Pro Ala Pro Pro Pro Ser Tyr Gly Ala Thr						
	530		535		540	
His Ser Thr Val Gln Pro Pro Arg Pro Pro Pro Pro Val Arg Asp Tyr						
	545		550		555	
						560

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 272 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT02
- (B) CLONE: 754412

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATTATATCA ATGTCTTAGA TATTGCTCCA AGGTATTCCA GTTTTCTCAT GGGAGCATCA	60
AGAGGATTTT CGAGCATAGC ACCTGTCATT GTACCCACTG TCAGTGGATT TCTTCTTAGT	120
CAGGACCCTG AGTTTGGGTG GAGGAATGTC TTCTTCTTGC TGTTTGCCGT TAACCTGTTA	180
GGACTACTCT TCTACCTCAT ATTTGGAGAA GCAGATGTCC AAGAATGGGC TAAAGAGAGA	240
AAACTCACTC GTTTATGAAG TTATCCCACC TT	272

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (B) CLONE: XLR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTTGATGCTC CCATGAGAAA ACTGG	25
-----------------------------	----

(2) INFORMATION FOR SEQ ID NO:7:

B1

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (B) CLONE: XLF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGGATTTTCG AGCATAGCAC CTGTC

25